

# Nederlandse Parels van Linked Data Toepassingen 2019

Gems of Dutch Linked Data Applications 2019



Platform Linked  
Data Nederland



PlatformLinkedData.nl



3	Voorwoord: Recente ontwikkelingen rond Linked Data
5	Inleiding PLDN
7	Linked Data vraagt open standaarden
9	Linked Aquo, de standaard voor waterbeheerders
12	Kadaster Dataplatform
14	De Nederlandse Bibliografie als Linked Open Data
16	Traceerbaarheid van producten in de chipindustrie
18	Digitaal Stedelijk waterbeheer
20	De Linked Data Voorziening bij de Rijksdienst voor het Cultureel Erfgoed
22	Een aan Wikidata gekoppelde thesaurus voor audiovisuele archieven
24	Wie, wat en waar in het onderwijs
26	Principes voor goed datamanagement en verantwoorde datawetenschap
28	Slimmer samenwerken in de spoorbranche met OTL en Linked Data
30	grlc: Eenvoudig en snel API's bouwen voor Linked Data
32	HortiCube: Linked Data voor de voedsel- en sierteeltindustrie
35	Lijst met afkortingen
36	Colofon



## Voorwoord

# Recente ontwikkelingen rond Linked Data

**De afgelopen jaren is er veel gebeurd op het gebied van Linked Data. Zo werkt TNO samen met diverse partijen in de domeinen agrifood, tuinbouw, industrie, overheid en energie aan oplossingen gebaseerd op Linked Data-technologie. Nadat de eerste versies van de W3C-standaarden RDF en OWL bijna 20 jaar geleden zijn vastgesteld, hebben de nodige ontwikkelingen plaatsgevonden.**

Het maken van RDF- en OWL-domeinmodellen heeft de afgelopen tien jaar voet aan de grond gekregen in diverse sectoren. Ook het gebruik van ontologie om semantiek van data formeel vast te leggen en daarmee interoperabiliteit tussen IT-systemen te realiseren is duidelijk in opmars. Vanuit deze historie gezien springen nu enkele recente ontwikkelingen in het oog.

De eerste ontwikkeling betreft de Shapes Constraint Language (SHACL), een van de meest recente specificaties die door W3C is opgesteld om semantiek van data nog beter vast te kunnen leggen. Deze specificatie biedt de mogelijkheid om condities te definiëren op data van een ontologie. Deze condities ('shapes')



omvatten een rijke verzameling aan taalconstructies zoals sh:minCount en sh:maxCount die de ‘vorm’ van de data afbakenen. De intentie van SHACL is dat een datamodel dat zou moeten voldoen aan een ontologie gevalideerd kan worden op basis van de ‘shapes’ zoals vastgelegd in die ontologie. Naast validatiemogelijkheden biedt SHACL ook een basis voor het bouwen van een user interface en codegeneratie op basis van de ‘shapes’ en hun condities.

De tweede ontwikkeling betreft de upgrade van Linked Open Data naar FAIR-data. Waar het bij Linked Open Data vooral gaat over het wereldwijd verbinden van data, gaat het bij FAIR om data die Findable (vindbaar), Accessible (toegankelijk), Interoperable (uitwisselbaar) en Reusable (herbruikbaar) is. Een belangrijk aspect van FAIR is dat data beschikbaar moet zijn onder duidelijk beschreven condities en licenties, met duidelijke referenties en goed beschreven metadata. Dus ook data, die om redenen van bijvoorbeeld privacy niet geheel open gepubliceerd kan worden, kan voldoen aan alle FAIR-principes. Daarmee voegt FAIR dus extra waarde toe aan Linked Data.

De derde ontwikkeling betreft de combinatie van Linked Data met rule-based reasoning. Alhoewel OWL-reasoners al beschikbaar zijn sinds ongeveer 2010, komt het gebruik ervan nu van de grond. Belangrijke drivers hiervoor zijn nieuwe talen zoals SWRL en SHACL die het mogelijk maken om regels te definiëren op instanties van concepten in ontologieën en daarover te redeneren in combinatie met OWL-reasoning-mogelijkheden op de concepten zelf. Aangezien redeneren een cruciale eigenschap is van intelligente systemen, ontstaan hiermee toepassingen van Linked Data binnen de kunstmatige

intelligentie. Bovendien vormen reasoners een ideale basis om uitleg te kunnen geven over de wijze waarop een kunstmatig-intelligent systeem tot een bepaalde conclusie komt, waardoor de gebruiker van het systeem het resultaat sneller als waardevol accepteert.

De vierde en laatste ontwikkeling betreft het toenemend gebruik van Linked Data in domeinen waarin het gebruik van ontologie en semantiek niet vanzelfsprekend is. Specifiek wordt in de telecommunicatiesector van oudsher gebruik gemaakt van communicatieprotocollen om connectie te maken en veilig datatransport te realiseren. Door de opmars van het aantal apparaten in huis, rondom de persoon en in de stad dat data produceert, neemt ook de behoefte toe om de semantiek van deze data goed vast te leggen. Een van de initiatieven die vanuit ETSI wordt ondersteund is de SAREF smart appliances reference ontology, die precies dat beoogt. Momenteel wordt SAREF verder toegespitst op de sectoren landbouw, energie en smart cities.

**Jack Verhoosel**  
TNO



## Inleiding

**Door verzamelingen van data te verbinden ontstaat er samenhang en kunnen er nieuwe inzichten tot stand komen. Het koppelen van data kan informatie opleveren die meer is dan de som der delen.**

Begin jaren negentig beseftte de overheid al hoe je data kon combineren om tot nieuwe inzichten te komen. Men zag toen al in dat je bijvoorbeeld steunfraude kon helpen opsporen door de database van uitkeringen te koppelen aan de voertuigenregistratie van de Rijksdienst van het Wegverkeer. Dit veroorzaakte ook toen al een discussie over privacy.

Nu, een kwart eeuw later, is heel veel data online toegankelijk en zijn de mogelijkheden om deze data te combineren bijna onbegrensd. Linked Data levert hier de technologie voor en sluit naadloos aan op de principes van het World Wide Web. De beginsels van Linked Data zijn:

- Benoem dingen met URI's.
- Zorg dat je via die URI's informatie kan vinden over de dingen die ze benoemen.
- Verbind URI's met andere URI's zodat de gecombineerde informatie nieuwe informatie oplevert.

Met Linked Data breng je dus samenhang aan in data. Door data te verbinden wint data aan betekenis en waarde.

Het idee achter Linked Data ontstond eind jaren negentig toen het World Wide Web

nog geen tien jaar bestond en volop groeide. Linked Data is inmiddels al twee decennia oud en al lang geen experimentele technologie meer waar alleen onderzoekscentra mee bezig zijn. Linked Data wordt echt gebruikt en biedt oplossingen voor actuele integratieproblemen.

Dit boekje geeft een aantal voorbeelden van het gebruik van Linked Data in de praktijk. Deze voorbeelden uit heel verschillende domeinen laten zien dat er met Linked Data oplossingen mogelijk worden die voorheen moeilijk te realiseren waren. Zo laat TNO zien hoe voedselproducenten gecombineerde open data over de landbouw en de voedselindustrie kunnen gebruiken om het aanbod beter bij de vraag te laten aansluiten. Kadaster maakt het mogelijk om basisregistraties te koppelen aan andere dataverzamelingen en de resultaten geografisch in kaart te brengen. NXP Semiconductors (chipindustrie) gebruikt Linked Data om individuele producten te kunnen volgen door de hele productie- en distributieketen heen. En zo laat dit boekje nog zeven andere praktijkvoorbeelden zien van Linked Data.



#### Platform Linked Data Nederland

Dit boekje is een publicatie van het Platform Linked Data Nederland, een netwerk waarin experts en belangstellenden kennis delen over Linked Data. Het platform bestaat uit mensen werkzaam in het bedrijfsleven, de overheid en kennisinstellingen. Er zijn nauwe banden met de wetenschap en met internationale organisaties. Het Big Data Value Center in Almere en het ECP bieden een thuis aan het Platform Linked Data Nederland, en leveren een omgeving waarin geëxperimenteerd kan worden met Linked Data. Het platform organiseert bijeenkomsten met verschillende reikwijdte, van kleinere expertbijeenkomsten tot grotere symposia voor een breder publiek.

#### Platform Linked Data Nederland

[platformlinkeddata.nl](https://platformlinkeddata.nl)

## Linked Data vraagt om open standaarden



#### Han Zuidweg

Forum Standaardisatie

**Linked data ontstaat door gegevens uit verschillende bronnen te combineren. Die bronnen kunnen heel divers zijn. Bronnen kunnen verspreid staan over het internet, door verschillende organisaties beheerd worden en uiteenlopende bestandsindelingen hebben. Alleen als je goede afspraken maakt over hoe je zulke uiteenlopende gegevensbronnen verbindt, kun je zinvolle en goed doorzoekbare koppelingen maken. Daarom zijn open standaarden essentieel.**



#### Forum Standaardisatie

Standaard Samenwerken

Linked Data is bijna synoniem met Resource Description Framework (RDF), de standaard waarmee je data verbindt in de vorm van 'onderwerp, gezegde, lijdend voorwerp'-tripels. RDF is een standaard van W3C, de organisatie die met open standaarden de fundamenteen legt voor het World Wide Web. Bouwend op RDF biedt de standaard SKOS een manier om een machineleesbaar kennissysteem te beschrijven. In een kennissysteem leg je vast hoe concepten met elkaar gerelateerd zijn. De Web Ontology Language-standaard (OWL) gaat nog een stapje verder en maakt het mogelijk om logische afleidingen te doen op relaties tussen concepten.

Linked Data bestaat niet zonder open standaarden.



# Linked Aquo, de standaard voor waterbeheerders

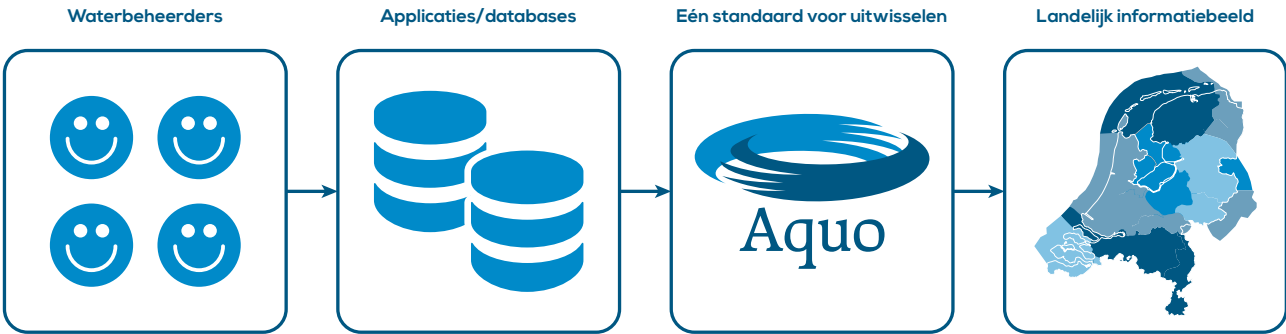
**Edwin Wisse**  
Informatiehuis Water

**De waterbeheerders van Nederland gebruiken de Aquo-standaard om gegevens over het waterbeheer op een uniforme manier uit te wisselen. De Aquo-standaard wordt onder andere gebruikt voor de uitwisseling van gegevens voor de Kaderrichtlijn Water, het Hoogwaterbeschermingsprogramma en het Inspire-programma.**

Het Informatiehuis Water (IHW) beheert de Aquo-standaard en zorgt dat deze beschikbaar is voor de gebruikers. Aquo omvat informatiemodellen voor waterkwaliteit, watersysteem en waterveiligheid, domeintabellen en het woordenboek.

De Aquo-standaard is vanaf 2019 als Linked Data beschikbaar. De Aquo-domeintabellen en het woordenboek zijn dan in één omgeving aan elkaar gekoppeld. Zo krijgt elke domeinwaarde een definitie. Met Linked Data kunnen begrippen in de Aquo-standaard eenvoudig gekoppeld worden aan begrippen in andere standaarden.

Het IHW heeft de Aquo-standaard als Linked Data beschikbaar gemaakt in een Dotwebstack-omgeving (Linked Data Theatre). Het mooie van Linked Data is dat het IHW niet bij elke domeinwaarde zelf een definitie



Recent heeft W3C een nieuwe Linked Data-standaard gepubliceerd, de Shapes Constraint Language (SHACL). Met SHACL kun je in een machineleesbare taal aangeven aan welke kwaliteitsvoorwaarden gegevens moeten voldoen, en wat er moet gebeuren als gegevens niet aan die voorwaarden voldoen. Zo kan je met SHACL bijvoorbeeld aangeven dat elke auto in een database met autogegevens een kenteken moet hebben. En dat er een foutboodschap moet worden getoond als een zoekopdracht in de database een auto oplevert die geen kenteken heeft.

Technisch gezien rust Linked Data op deze relatief eenvoudige open standaarden. Dat wil niet zeggen dat het publiceren van Linked Data gemakkelijk is. Er is veel denkwerk voor nodig om dataverzamelingen op een betekenisvolle manier te verbinden. Kwalitatief goede Linked Data maken is arbeidsintensief. Daarom is Linked Data bij de overheid nog geen gemeengoed, maar wel in opkomst.

De Linked Data-standaarden RDF, OWL, SKOS en SHACL staan op de lijst open standaarden die het Forum Standaardisatie aanbeveelt. De Nederlandse overheid stelde het Forum Standaardisatie ruim tien jaar geleden

in om het gebruik van open standaarden in de publieke sector te stimuleren. Deskundigen uit de overheid, het bedrijfsleven en de wetenschap vormen samen het Forum Standaardisatie. Zo vinden we er onder andere vertegenwoordigers van de ministeries van Economische Zaken en Binnenlandse Zaken en Koninkrijksrelaties, het Interprovinciaal Overleg, VNG Realisatie, het Centraal Bureau voor de Statistiek, VNO-NCW, Geonovum en de Universiteit van Eindhoven.

Het Forum Standaardisatie heeft de Linked Data-standaarden ook namens Nederland aangemeld bij het European Multi-Stakeholders Platform on ICT Standardisation. Deze expertgroep van de Europese Commissie identificeert standaarden die bij aanbestedingen in Europa uitgevraagd kunnen worden.

Het Forum Standaardisatie steunt het Platform Linked Data Nederland (PLDN) omdat overheidsorganisaties het belang van Linked Open Data inzien. Samen met het PLDN wil het Forum Standaardisatie ervoor zorgen dat de overheid haar data publiceert in formaten die voldoen aan open standaarden en aan de wetgeving rondom toegankelijke, herbruikbare overheidsinformatie.



hoeft te geven, maar ook kan verwijzen naar al bestaande definities in andere woordenboeken.

Het IHW koppelt de Aquo-standaard op objectniveau met andere standaarden. Hierbij worden relaties gelegd met begrippen in Aquo, die overeenkomen met begrippen in het Gegevenswoordenboek Stedelijk Water (GWSW) van de Stichting RIONED. Wanneer er een goede externe definitie bestaat voor een begrip zoals 'gemaal', dan verwijst het IHW hiervoor door naar het GWSW. Op dezelfde manier is de Aquo-standaard met Linked Data-technieken verbonden met Seadatanet, een standaard die voor maritieme informatie gebruikt wordt. Het feit dat zowel de GWSW- als Seadatanet-woordenboeken als



Linked Data beschikbaar zijn, maakt flexibele koppeling van begrippen tussen woordenboeken mogelijk. Dit is een hele stap vooruit.

Het omzetten van Aquo naar Linked Data heeft het IHW beter inzicht gegeven in de inhoud van Aquo en in de mogelijkheden om de standaard beter te richten op de essentie van het watermanagement. Met Linked Data wordt het makkelijker om standaarden te verbinden. Het IHW koppelt de Aquo-standaard in de toekomst aan nog meer standaarden, ook internationale. Ook wil het IHW Linked Data gebruiken om de relaties binnen de Aquo-standaard tegen het licht te houden, dus definities van begrippen bij domeinwaarden en omgekeerd.



Linked Data zorgt voor eenduidige begrippen in standaarden voor het waterbeheer.



# Kadaster Dataplatform

Erwin Folmer  
Kadaster

**Het Kadaster heeft als ambitie om betrouwbare data beschikbaar te stellen waarmee maatschappelijke en economische waarde wordt gecreëerd kan worden. Daarnaast is er de ambitie om het 'Geoplatform van Nederland' te bieden waarmee iedereen met geografische informatie altijd en overal aan de slag kan.**

Deze ambitie wordt grotendeels ingevuld met de Publieke Dienstverlening op Kaart (PDOK), een gezamenlijk initiatief van de Nederlandse overheid waarvan het beheer bij het Kadaster is ondergebracht. PDOK biedt al vijf jaar een hoogwaardig dataplatform waarop meer dan 100 datasets van verschillende overheden worden aangeboden in geografische formaten.

Het Kadaster Dataplatform is een van de diensten die via het PDOK beschikbaar zijn. Met de ontwikkeling van het Kadaster Dataplatform zet Kadaster de trend voor het dataplatform van de toekomst gebaseerd op Linked Data, dat invulling moet geven aan de ambitie van 'betrouwbare data voor iedereen'.

Voor betrouwbare data gelden deze principes:

1. Houd data bij de bron: voorkom zoveel mogelijk dat gebruikers data kopiëren, want hierdoor wordt de data minder betrouwbaar.
2. Geef semantiek bij de data: hiermee voorkom je dat data verkeerd geïnterpreteerd wordt.
3. Biedt data in samenhang aan: zowel de betrouwbaarheid als de toegevoegde waarde stijgt als de data uit verschillende datasets met elkaar verbonden worden.

Een ander belangrijk aspect is 'data voor iedereen'. Een 'one size fits all'-aanpak werkt niet, zodat data in verschillende formats moet worden aangeboden.

## Basisregistraties: betrouwbare data voor iedereen dankzij Linked Data



kadaster



Alleen met Linked Data kunnen deze principes waargemaakt worden. Een alternatief is er eigenlijk niet. Toch is Linked Data alleen niet voldoende. Het Kadaster Dataplatform publiceert data ook via REST-API's. Hiermee kunnen informatieproducten op maat gemaakt worden en bruikbaar voor een brede groep van webdevelopers. Op de achtergrond worden de API's gevoed met Linked Data.

Het Kadaster Dataplatform is al geruime tijd operationeel, en biedt meerdere datasets aan, waaronder Basisregistratie Adressen en Gebouwen (BAG) en de Basisregistratie Topografie (BRT). Doordat het platform ook via PDOK wordt aangeboden en de mogelijkheid bestaat om eigen branding op een dataset te doen (bijvoorbeeld met eigen webadressen of URIs), biedt het Kadaster Dataplatform ook een thuis voor andere datasets zoals de Wijken en Buurten van het Centraal Bureau voor Statistiek. Voor de komende tijd staan

er vele nieuwe datasets gepland. Ook is het Kadaster Dataplatform een onmisbare schakel geworden voor Digitaal Stelsel Omgevingswet.

Naast het publiceren van Linked Data sets en API's biedt het Kadaster Dataplatform hulpmiddelen om het gebruik van deze data te stimuleren. Met Data Stories kan een verhaal gemaakt worden aan de hand van vooraf gedefinieerde SPARQL-queries. Data Stories laten de kracht van het live bevragen van data goed zien. Met FacetCheck hebben we een filtergebaseerde browser op data, een mooie manier om door de data te grasduinen.

Experimenteel onderzoekt het Kadaster ook de integratie met business intelligence tools en andere Data Science-toepassingen. Deze ontwikkelingen doet het Kadaster zoveel mogelijk in openheid op de labs-omgeving.



# De Nederlandse Bibliografie als Linked Open Data

Enno Meijers

Koninklijke Bibliotheek

**De Koninklijke Bibliotheek (KB) heeft als missie om vanuit de kracht van het geschreven woord bij te dragen aan een slimmer, vaardiger en creatiever Nederland. We doen dit als netwerkorganisatie samen met vele partners in het domein van de openbare bibliotheken, het erfgoed en de wetenschap. Als nationale bibliotheek maakt de KB de bibliotheekcollectie Nederland zichtbaar, bruikbaar en houdbaar.**

Voor het beter zichtbaar maken van het gezamenlijk bezit speelt de beschrijving van de collecties een belangrijke rol. Zo'n beschrijving helpt om inzichtelijk te maken wie welke informatie heeft en hoe deze te benaderen is. Al vroeg in de vorige eeuw werkten bibliotheken aan het systematisch beschrijven van hun collecties, vastgelegd in kaartcatalogi. Vanaf de jaren 80 is men deze informatie gaan digitaliseren en ontstonden de bibliotheekcatalogi zoals we die nu kennen.

De oorspronkelijke werkwijze is nog altijd terug te vinden in de huidige informatiesystemen zoals in het meta-data-format en de manier waarop gegevens gedeeld worden. Nationale bibliotheken verzorgen, vaak in samenwerkingsverbanden, bibliografische beschrijvingen voor de publicaties in hun werkgebied. De overige bibliotheken kopiëren deze vervolgens naar hun eigen systemen om daarmee hun eigen diensten te bouwen. Vaak ontstaat er zo ook nieuwe informatie in deze decentrale systemen.

Linked Data maakt de Nederlandse Bibliografie toegankelijk bij de bron.

**KB** Koninklijke Bibliotheek  
Nationale bibliotheek van Nederland

Het resultaat van de huidige manier van werken is dat er veel kopieën ontstaan waarbij het moeilijk is om de bron te herkennen en een overzicht te krijgen van alle beschikbare informatie. Soms lossen bibliotheken dit op door samen te werken binnen netwerken in een gezamenlijk systeem. Dit vraagt om veel samenwerking en coördinatie en het resultaat is vaak niet optimaal. Extra kosten voor het in standhouden van het gezamenlijke systeem en afspraken over het gebruik van de informatie maken dit in de praktijk nog lastiger.

Linked Data biedt een oplossing voor dit probleem. Door deze informatie niet langer te kopiëren naar andere bibliotheeksystemen maar dit op een bruikbare manier direct op het web te publiceren, kunnen de gegevens met Linked Data-technieken gekoppeld worden. Bibliotheken kunnen nu volstaan met het verwijzen naar de gepubliceerde beschrijvingen. Wanneer ze vervolgens ook hun eigen verrijkingen hieraan koppelen en open publiceren, ontstaat er een web van bibliografische informatie. Het voordeel van deze werkwijze is dat de bron duidelijk blijft, de informatie direct beschikbaar is en iedereen hier met eenvoudige middelen van kan profiteren.

Als een eerste stap in deze richting publiceert de KB sinds eind 2017 de Nationale Bibliografie als Linked Open Data. Beschikbaar voor iedereen om naar te verwijzen en te gebruiken. Hoewel het nog gaat om een bètaversie is de dienst wel ontworpen als een duurzame informatiebron waar anderen op kunnen bouwen.



In het afgelopen jaar hebben Wikidata, DBpedia, Europeana en diverse erfgoedinstellingen al gebruik gemaakt van deze informatie. Ook individuele data-onderzoekers weten de weg inmiddels te vinden.

In de komende jaren zal de KB steeds meer bronnen als Linked Open Data beschikbaar gaan maken waaronder de Nederlandse Thesaurus van Auteursnamen en diverse andere thesauri. In de eerste plaats gaat dit om bronnen die we zelf beheren. De KB werkt zo aan betere zichtbaarheid van de alle publiek gefinancierde bibliotheekcollecties.

[data.bibliotheken.nl](http://data.bibliotheken.nl)



# Traceerbaarheid van producten in de chipindustrie

John Walker  
Semaku B.V.

Onno Buijs  
NXP Semiconductors B.V.

Voor fabrikanten van chips is traceerbaarheid van producten cruciaal. Als zich een probleem met een product voordoet, is het van groot belang om het defecte onderdeel te identificeren, het achterliggende probleem te vinden en de getroffen delen veilig te stellen. NXP Semiconductors traceert daarom elke partij producten door het hele productie- en leveringsproces heen.

De productie- en leveringsdata wordt over verschillende systemen verspreid bewaard. De meeste van deze systemen zijn relationele databases. NXP Semiconductors gebruikt nu Linked Data-technologie om de data in de verschillende systemen te koppelen. De gecombineerde data kan worden doorzocht en geanalyseerd via SPARQL-queries en JSON-LD-API's.

Het project is in verschillende fases tot stand gekomen. In eerste instantie bouwde NXP Semiconductors een RDF graph voor de belangrijkste database in het rapporteringsproces, een complexe database met verschillende tabellen en meer dan 100 miljoen records in de hoofdtabel. Het semantisch model bleek te complex voor een werkbare Linked Data-representatie, en ook de prestaties vielen tegen door de zware bevraging van de database die de SPARQL-queries teweegbrachten.

Maar de resultaten waren voldoende bemoedigend om te werken aan een efficiëntere en meer schaalbare Linked Data-structuur. In het tweede ontwerp betrok NXP Semiconductors de gebruikers vanaf het begin waardoor de Linked Data beter aansloot bij de behoeften van de professionals die klachten van klanten moeten afhandelen en kwaliteitsproblemen moeten opsporen en verhelpen. SPARQL voor tabellen, ook bekend als Tarql, bood een oplossing voor het schaalprobleem. Ook de userinterface werd opnieuw ontwikkeld, en zo ontstond de Global Trace Portal 2.0.

Dankzij de Linked Data-aanpak laat de Global Trace Portal 2.0 het ook toe om individuele producten te volgen. Dit is een belangrijke verbetering ten opzichte van het

oorspronkelijke systeem dat alleen partijen volgde. In een partij kunnen wel duizenden producten zitten, waardoor een product met een probleem er moeilijk uit te lichten is. De Global Trace Portal 2.0 combineert de gegevens over partijen met gegevens over individuele producten, waardoor er op individuele productcodes gezocht kan worden.

Bij testen blijkt de Global Trace Portal 2.0 efficiënt en schaalbaar. Gebruikersacceptatietesten laten zien dat gebruikers het nieuwe interactieve ontwerp en gebruiks-



vriendelijkheid van de applicatie waarderen. Het portaal is negen maanden in gebruik met inmiddels 100 actieve gebruikers. Het portaal functioneert goed. Wel wordt steeds duidelijker dat de koppeling met databronnen dynamischer moet worden, want er komen meer en meer aanvragen voor extra systeemkoppelingen. Dit brengt ook weer schaalbaarheidsproblemen met zich mee. Caching van data moet hiervoor een oplossing bieden. Daar wordt nu hard aan gewerkt. Total Traceability is een belangrijk onderdeel van NXP's Total Quality-programma!

Dankzij Linked Data kunnen producten door de hele productie- en distributieketen getraceerd worden.

[semaku.com/qualityassurance.html](https://semaku.com/qualityassurance.html) | [nxp.com](https://nxp.com)



# Digitaal stedelijk waterbeheer



Eric Oosterom  
Stichting Rioned

**Stedelijk waterbeheer is een kerntaak van gemeenten om onze maatschappij gezond, leefbaar en veilig te houden. De gemeenten hebben van oudsher uitgebreide, maar versnipperde bestanden opgebouwd met objectkenmerken, conditiebepalingen en meet- en regelgegevens. De interoperabiliteit van die gegevens is gebrekkig, de datasilo's zijn vaak nauw verbonden met specifieke applicaties. Dat is de belangrijke motivatie voor het ontwikkelen van het Gegevenswoordenboek Stedelijk Water (GWSW).**

Het GWSW, een op Linked Data gefundeerde semantisch rijke ontologie, datastructuur en uitwisselmethodiek, vormt de basis voor een betere beschikbaarheid, bruikbaarheid en kwaliteit van de datasets voor het stedelijk waterbeheer. Administratieve systemen, productcatalogi, (contract)eisenbibliotheken en beheermethodieken kunnen de GWSW-datasets gebruiken. Met het GWSW ontstaat een 'digitale tweeling' van het fysieke systeem, zodat de gehele levenscyclus van idee via ontwerp, realisatie en beheer tot sloop (en hergebruik) digitaal gefaciliteerd kan worden.

De ontwikkeling en invoering van het GWSW is een proces van lange adem. Vanaf 2015 komen er al toepassingsgerichte modules beschikbaar voor geo-informatie (ligging en kenmerken), hydraulisch modelleren en beheren (aanleggen, reinigen, inspecteren, vervangen, renoveren en repareren). Daar gaat nu data over gemalen, persleidingen, meldingen, meten en monitoren bij komen.

De belangrijkste leveranciers van beheerssoftware hebben inmiddels import- en exportfuncties voor GWSW-datasets ingebouwd, waardoor de gemeenten hun beheergegevens kunnen omzetten naar GWSW-conforme datasets, die kunnen valideren, en vervolgens in de cloud kunnen plaatsen. Zo'n 50 gemeenten hebben dit medio 2018 gedaan en het aantal groeit snel. Zo ontstaat een landelijke datavoorziening met gemeentelijke datasets. Nu is dit nog een kopie van de native datasets, maar in de toekomst blijft de data bij de bron in de cloud.

Linked Data en een verbindend woordenboek bouwen een landelijke voorziening van gemeentelijke datasets voor waterbeheer.

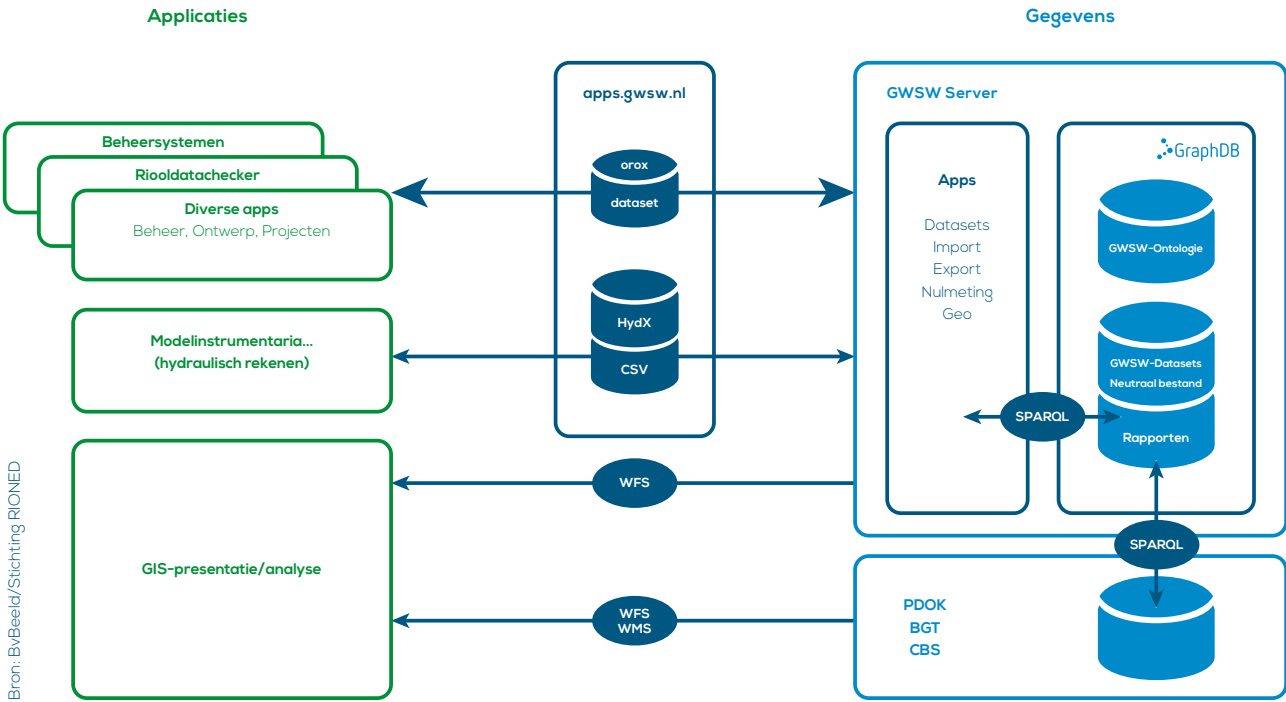
data.gwswnl

Stichting RIONED ontwikkelt en onderhoudt de GWSW-ontologie. Aan die datastructuur wordt veel aandacht besteed, de basis daarvoor wordt gevormd door normen (bijv. EN13508-2 voor rioolinspectie), eisen, specificaties en de kennis van experts in diverse werkgroepen.

Het GWSW-OroX-protocol op basis van RDF, RDFS, OWL-2 en SKOS definieert een serie vaste relaties ('predicates') die niet alleen de concepten in het GWSW maar ook de gemeentelijke datasets beschrijven. Die relaties zijn specifiek genoeg om processen (activiteiten met in- en uitvoer), decompositie, soortenboom, geometrie en topologie te beschrijven.

De GWSW Server, voornamelijk in proefopstelling bij Stichting RIONED, faciliteert de upload van gemeentelijke datasets en de basistoepassingen. Belangrijke voorbeelden daarvan zijn de geografische publicaties conform het WFS-protocol (GML) en de basis-kwaliteitsmeting van de gegevens.

De datasets op de GWSW Server zijn via een SPARQL-endpoint eenvoudig te combineren met andere Linked Data binnen het semantisch web. Recent zijn geslaagde proefnemingen gedaan in samenwerking met Kadaster en gepubliceerd via PDOK (www.pdok.nl).



Bron: BvBeeld/Stichting RIONED

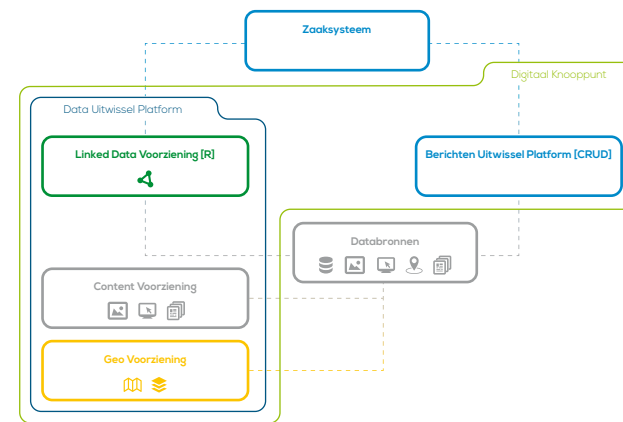


# De Linked Data Voorziening bij de Rijksdienst voor het Cultureel Erfgoed

**Joop van der Heijden**  
Rijksdienst voor het  
Cultureel Erfgoed

De Rijksdienst voor het Cultureel Erfgoed (RCE) stelt informatie over cultureel erfgoed beschikbaar aan het publiek. De manier waarop dit gebeurt kan nog verbeterd worden. Informatie moet beschikbaarder, bruikbaarder, bestendiger en makkelijker te combineren worden.

Een voorbeeld is het Nationaal Monumenten Portaal dat gegevens publiceert over rijksmonumenten. Nu levert het RCE periodiek een berg gegevens aan de beheerder van het Nationaal Monumenten Portaal, die deze gegevens dan op het portaal plaatst. Na een paar weken zijn de gegevens niet meer actueel en moet de RCE een bijgewerkte set gegevens aanleveren. Dit kan veel beter door de data bij de bron te laten en Linked Data-technieken te gebruiken.



De RCE heeft een Data Uitwissel Platform in het leven geroepen waarmee alle RCE-informatie volledig doorzoekbaar wordt. Een Linked Data Voorziening maakt onderdeel uit van het platform, dat daardoor een duurzame, efficiënte, flexibele en integrale oplossing kan bieden voor informatie-uitwisseling. De Linked Data Voorziening biedt alle gegevens van het RCE aan als Linked Data en maakt ze via een API beschikbaar.

Voor de Linked Data Voorziening gebruikt de RCE een bi-modale architectuur die bestaat uit een robuust platform en een flexibel platform. Het robuuste platform werd ontwikkeld in samenwerking met de Publieke Dienstverlening Op Kaart (PDOK). Het flexibele platform is gebaseerd op de PoolParty Semantic Integrator van de Semantic Web Company.

Aan de basis van de Linked Data Voorziening ligt de Cultureel Erfgoed Ontlogie (CEO). Deze ontlogie legt alle termen vast die nodig zijn om cultureel erfgoed-objecten en de kennis daar omheen te beschrijven. De CEO heeft verbindingen met de Basisregistratie Kadaster (BRK) en de Overheid Web Metadata Standaard (OWMS).

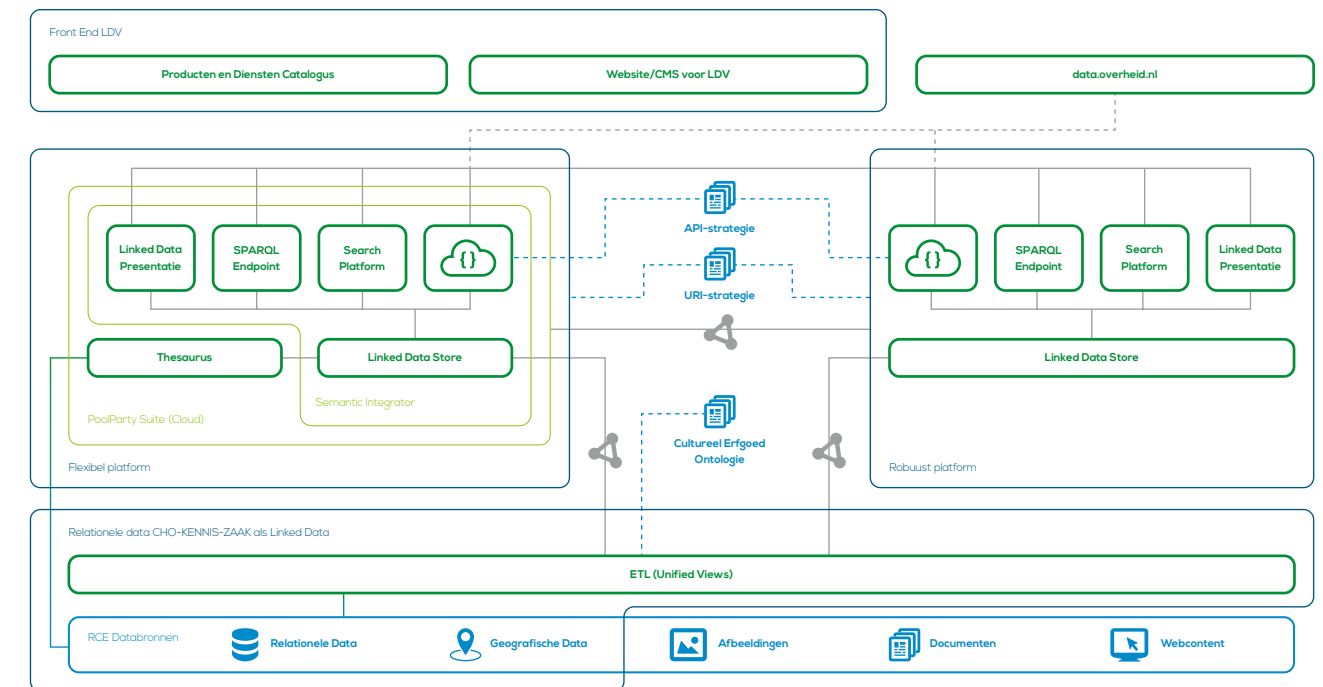
De Linked Data Voorziening werd aanvankelijk als een puur ICT-project gezien, maar gaandeweg werden ook data-eigenaren steeds meer betrokken. Dit bleek erg belangrijk om de transitie van de gegevensbronnen naar Linked Data te realiseren. De data-eigenaren brachten ook

cultureelerfgoed.nl

De Linked Data Voorziening  
houdt doorzoekbare en  
bruikbare data over  
nationaal erfgoed  
bij de bron.

ideeën in om data op nieuwe manieren te verbinden. Zo werd er gekeken naar aanpassingen van het proces om gegevens over adressen en percelen te koppelen.

De Linked Data Voorziening gaat eind 2018 live. In eerste instantie gebruikt de RCE Linked Data om partners zoals het Netwerk Digitaal Erfgoed en Digitaal Stelsel Omgevingswet te voorzien van gegevens. Daarna wil het RCE meer gegevensverzamelingen publiceren, bijvoorbeeld alle afbeeldingen. De RCE wil Linked Data uiteindelijk ook gebruiken om de eigen processen te ondersteunen.





# Een aan Wikidata gekoppelde thesaurus voor audiovisuele archieven

Johan Oomen

Nederlands Instituut  
voor Beeld en Geluid

**Beeld en Geluid beheert sinds jaar en dag een eigen thesaurus om de eigen collectie van audiovisueel erfgoed te beschrijven. Deze Gemeenschappelijke Thesaurus voor Audiovisuele Archieven (GTAA) wordt ook door andere collectiebeherende instellingen gebruikt.**

In de GTAA komen ruim 130.000 persoonsnamen voor. Door deze persoonsnamen te koppelen aan andere databronnen worden er heel nieuwe toepassingen mogelijk. Zo wordt de auteursrechtelijke status van een audiovisueel werk duidelijk als de sterfdatum van de auteur bekend is. Zeventig jaar na het overlijden van een maker komt materiaal in het publieke domein. Jaarlijks worden de medewerkers van Beeld en Geluid erop gewezen voor welke werken de auteursrechten op deze manier verlopen.

Om dit te bereiken koppelde Beeld en Geluid de GTAA aan Wikidata (wikidata.org) met Linked Data-technieken. In Wikidata bevatten personen naast namen vaak aanvullende informatie die niet, of in ieder geval niet gestructureerd beschikbaar is binnen de bronnen van Beeld en Geluid. Denk daarbij bijvoorbeeld aan geboorte- en sterfdaten, beroepen, geslacht en lidmaatschap van politieke partijen. Behalve uitgebreider, is de informatie op Wikidata over personen ook actueler doordat een grote community van vrijwilligers de data continu bijhoudt.

Door zoektechnologie van Spinqe heeft Beeld en Geluid zoveel mogelijk geautomatiseerde overeenkomsten gemaakt tussen personen in de GTAA en Wikidata. Daarbij werd gebruik gemaakt van de zogenaamde 'scope notes' uit de thesaurus, die zeer beknopte informatie bevatten over personen. Maar ook werd informatie uit de catalogus gehaald: woorden die vaak voorkwamen in combinatie met de betreffende persoon, of rollen die de persoon speelde in diverse producties.

Deze overeenkomsten werden aan de Wikidata-community aangeboden die er binnen enkele maanden ruim 30.000 heeft goedgekeurd. Dit aantal groeit nog dagelijks. Met deze verrijkte data kan Beeld en Geluid nu heel nieuwe dingen doen zoals het afleiden van het auteursrecht uit de sterfdatum van de maker.

Ook biedt de gekoppelde data geavanceerde zoekmogelijkheden. Door middel van SPARQL-queries kan op Wikidata bijvoorbeeld een lijst uitgevraagd worden van vrouwen die lid van de VVD zijn, als beroep 'politica' hebben en geboren zijn in de jaren 80. Dergelijke zoekvragen kunnen onderzoekers helpen bij het vinden van het juiste materiaal voor hun onderzoek.

Tenslotte vereenvoudigt de koppeling ook het onderhoud aan de GTAA. Het invoeren en beheren van de eerdergenoemde 'scope notes' is arbeidsintensief en foutgevoelig. Door te linken naar Wikidata heeft Beeld en Geluid deze werkdruk kunnen verlagen.

De komende jaren verwacht Beeld en Geluid ook eigen datadonaties aan Wikidata te kunnen maken van veel voorkomende personen uit haar media-archief. Ook zal worden gekeken naar de mogelijkheid van het koppelen van andere aspecten van de GTAA. Wikidata bevat in veel gevallen meerdere thesauri of databases bij personen. Wikidata kan dus als tussenlaag dienen om de collectie van Beeld en Geluid met nog weer andere databronnen te verbinden.

BEELD EN GELUID



Dankzij  
Linked Data  
kan je in  
audiovisueel  
erfgoed  
gedetailleerd  
en up-to-  
date zoeken  
op personen.

[gtaa.beeldengeluid.nl](http://gtaa.beeldengeluid.nl)



# Wie, wat en waar in het onderwijs

**Gerald Groot Roessink**  
Dienst Uitvoering  
Onderwijs

**De Dienst Uitvoering Onderwijs (DUO) voert wettelijke taken uit op het gebied van bekostiging, toezicht en verantwoording van het onderwijs in Nederland. DUO beheert het register van instellingen die actief zijn in het formele onderwijs met gegevens over de opleidingen die ze mogen geven. DUO publiceert dit register al jaren als open data.**

Wat DUO publiceert en beheert is zeker niet de hele werkelijkheid. In het register ontbreken de nuances van de werkelijkheid van de onderwijsinstellingen. Bijvoorbeeld een plusprofiel bij een HAVO-opleiding, of een leslocatie die wel bestaat maar administratief onzichtbaar is. Dit soort nuances zijn niet relevant voor DUO maar wel voor de instelling zelf, de studenten, en anderen zoals de Inspectie en het College voor Toetsen en Examens.

Daarom werkt DUO samen met de onderwijssector aan de Registratie Instellingen en Opleidingen (RIO). RIO houdt vanaf het ontwerp rekening met de verschillende invalshoeken. Linked Data leent zich bij uitstek voor dit soort samenwerking die organisaties overstijgt.

Sinds 2017 experimenteert DUO daarom met Linked Data. DUO gebruikt daarbij een Dotwebstack-omgeving (voorheen Linked Data Theatre) en de online Linked Data-voorziening van het Kadaster in de cloud. De Linked Data kan worden bevestigd via een SPARQL-endpoint maar ook met REST API's.

RIO geeft met Linked Data meer detail over wie wat waar aan onderwijs aanbiedt.

Terwijl er nog aan RIO wordt gewerkt ontsluit DUO de registergegevens over het wie, wat en waar in het onderwijs nu al als Linked Data. Door deze data te koppelen met data van de onderwijsinstellingen ontstaat er meer nuance. Vanaf 2019 zal de DUO Linked Data-component van RIO in productie zijn.

Daarnaast gaat DUO de RIO koppelen aan de Basisadministratie Adressen en Gebouwen (BAG) voor locaties en de stelselcatalogus voor het beschrijven van metadata van de basisregistraties. En in het onderwijs zijn plannen om onderwijsstructuren als Linked Data te publiceren, die als bron gekoppeld kunnen worden aan de RIO.

[www.rio-onderwijs.nl](http://www.rio-onderwijs.nl)





# Principes voor goed datamanagement en verantwoorde datawetenschap

Michel Dumontier  
Institute of  
Data Science  
Maastricht University

**De samenleving digitaliseert, en dat betekent dat er steeds meer gegevens worden ontsloten. Steeds meer onderzoeken maken gebruik van online gepubliceerde gegevens. Goed gegevensbeheer en verantwoordelijke gegevenswetenschap zijn vereisten voor een doordacht en veilig gebruik van online data. Slecht beheer en onverantwoord gebruik van gegevens kunnen veel risico's met zich mee brengen voor de samenleving.**

Het programma 'The Digital Society' van de Vereniging van Universiteiten in Nederland (VSNU) heeft vooraanstaande onderzoekers van alle veertien Nederlandse universiteiten samengebracht om te werken aan de uitdagingen die de digitale samenleving met zich meebrengt. De VSNU heeft twee beginselen beschreven voor de rol van online data in de digitale samenleving: FAIR voor goed gegevensbeheer en FACT voor verantwoorde gegevenswetenschap.

FAIR staat voor Findable, Accessible, Interoperable and Reusable data. De FAIR-principes van gegevensbeheer zijn ontwikkeld om onderzoekers te begeleiden bij het verbeteren van de herbruikbaarheid van hun gegevens door het effectief beheer daarvan. Goed databeheer helpt bij het optimaliseren van het rendement uit investeringen in onderzoek. Goed databeheer levert hoogwaardige digitale activa op die ook in de toekomst goed vindbaar en herbruikbaar blijven. Naast een juiste manier van verzamelen, annoteren en archiveren omvat goed gegevensbeheer daarom ook het begrip 'langdurige zorg'.

## Nederland loopt voorop met principes voor goed databeheer en verantwoorde data science.

Het FAIRmetrics, een samenwerkingsverband tussen het Dutch Techcenter for Life Sciences (DTL), de GOFAIR-stichting en het Institute of Data Science van Maastricht University, ontwikkelt oplossingen en instructies om FAIR-principes praktisch toe te passen. De FAIRmetrics-groep ontwikkelt een maatstaf waarmee kan worden bepaald in hoeverre gegevensbeheer FAIR is. Dit helpt organisaties om te bepalen waar ze staan in het toepas-

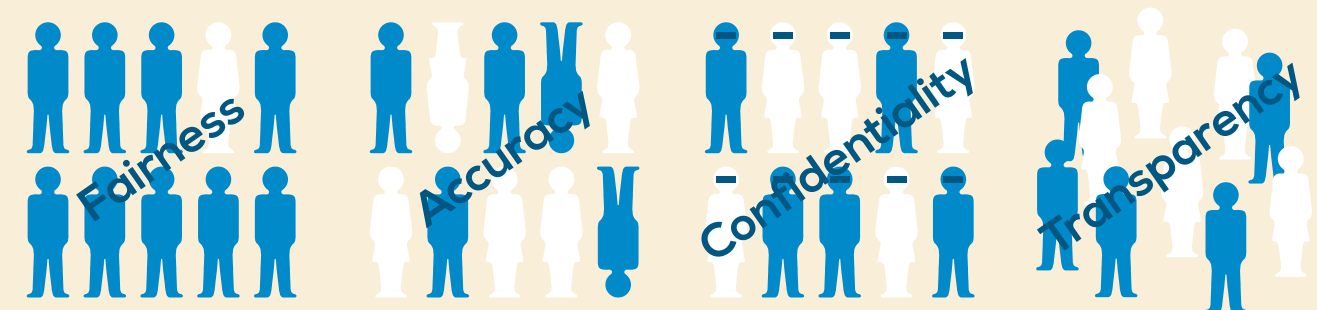
sen van FAIR-databeheer. Organisaties zoals ZonMw en de Hartstichting vragen onderzoekers nu al om het proces van gegevensmanagement te documenteren en de FAIR-principes vanaf het begin toe te passen. Het lijkt eenvoudig om gegevens te verzamelen in een FAIR-manier. Vaak hebben onderzoekers echter geen idee hoe ze dat zouden moeten doen. Daarom heeft de FAIRmetrics zo'n belangrijke rol bij het bouwen van een digitale samenleving.

FACT staat voor Fair, Accurate, Confidential and Transparent, en is van toepassing op gegevenswetenschap. De FACT-beginselen wijzen onderzoekers en studenten op het verantwoordelijk toepassen van gegevenswetenschap. Data-onderzoekstechnieken, -infrastructuren en -benaderingen moeten voorbedacht verantwoordelijk worden gemaakt.

Het Responsible Data Science-consortium (RDS) is opgericht om misbruik van gegevens te voorkomen. Het RDS-consortium brengt een multidisciplinaire groep experts en ideeën bij elkaar die een wetenschappelijke richting zoekt om gegevens op een veilige en verantwoorde manier te gebruiken. De resultaten van het RDS-programma omvatten nieuwe data-onderzoekconcepten, nieuwe analyse- en datamanagementtechnieken, krachtige softwaretools en -infrastructuren en RDS-educatie en -training. Organisaties zoals het Rathenau Instituut, het Centraal Bureau voor de Statistiek (CBS) en de Nederlandse Wetenschappelijke Raad voor het Regeringsbeleid (WRR) ondersteunen dit RDS-initiatief.

De FAIR- en FACT-principes moeten ook buiten de wetenschappelijke gemeenschap hun toepassing vinden. Hiervoor wordt nu al onderwijsmateriaal en infrastructuur ontwikkeld.

### FACT: Responsible Data Science



[vsnu.nl/en\\_GB/digitalsociety](https://vsnu.nl/en_GB/digitalsociety) | [responsibledatascience.org](https://responsibledatascience.org) | [maastrichtuniversity.nl/research/institute-data-science](https://maastrichtuniversity.nl/research/institute-data-science)

# Slimmer samenwerken in de spoorbranche met OTL en Linked Data

**ProRail**

René Wubbels  
ProRail

In de spoorbranche werkt ProRail samen met marktpartijen om één totaalproduct te leveren: een werkende infrastructuur. Een heldere communicatie en uitwisseling van informatie is daarbij onontbeerlijk. Spraakverwarring ontstaat als er voor hetzelfde object meerdere woorden bestaan zoals een 'laagsein', 'dwergsein' en 'hoofdsein'. Vakmensen kunnen hiermee nog wel uit de voeten, maar voor ICT-systemen is gebrek aan eenduidigheid een groot probleem.

De Object Type Bibliotheek van ProRail (OTL: Object Type Library) is een mijlpaal in de ontwikkeling van informatie-standaarden ten behoeve van de samenwerking in de spoorbranche. Met de OTL heeft ProRail de weg ingeslagen naar een gedeeld spoorbeeld, een sterk verbonden keten, een versneld proces en betrouwbare data. Mede door de OTL is het mogelijk dat infra-object-data binnen ProRail een digitale kopie is van de infrastructuur buiten: 'binnen = buiten'.

De OTL maakt het integraal informatiemodel (IIM) van ProRail toegankelijk. Het IIM is een ontologie die de structuur, begrippen, definities, taxonomie, eigenschappen en relaties van spoorse objecttypen generiek beschrijft. De OTL maakt de kennis die de IIM-ontologie 'onder de motorkap' vastlegt, praktisch bruikbaar voor het maken, wijzigen en uitwisselen van informatie gedurende de hele levenscyclus van een infrastructuurobject. Hiermee is de OTL meer dan een informatiestandaard; het is ook een kennisbank en validiteitschecker.

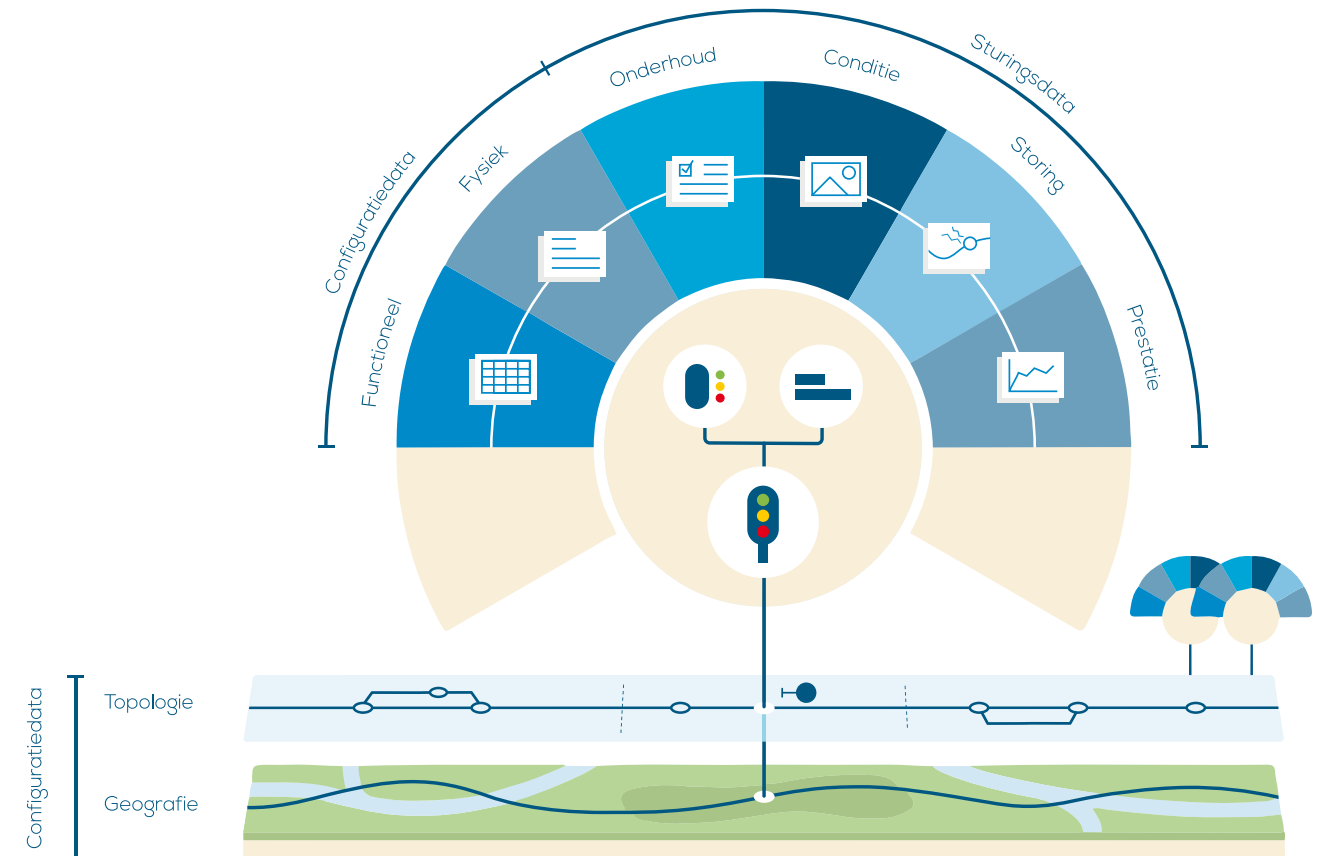
## De Object Type Library van ProRail komt binnenkort ook als Linked Data beschikbaar.

Dankzij de OTL zijn begrippen en definities gedurende alle fasen van het bouwwerk eenduidig gedefinieerd. ICT-systemen voor de bouw en het beheer van de spoorinfrastructuur werken door gebruik te maken van de OTL allemaal met dezelfde informatie en kennis. De OTL biedt daarnaast een platform voor het stellen van vragen of het indienen van wijzigingsverzoeken. Zo werken we als spoorsector samen aan de verbetering van het fundament onder de informatiehuishouding voor het spoor.

otl.prorail.nl

In oktober 2018 ging de OTL live. De website <https://otl.prorail.nl/> is openbaar en kan door iedereen geraadpleegd worden. De gebruiker kan objecttypen van de spoorinfrastructuur zoeken, bekijken en exporteren in spreadsheet-formaat. De website maakt thans gebruik van conventionele web-technologie, dus html en XML. Binnenkort komt de OTL ook beschikbaar als Linked Data, gebruik makend van RDF en OWL. Zo wordt de

OTL benaderbaar als een endpoint en kan via SPARQL worden bevraagd. Informatie is dan vindbaar op het web in een vaste structuur, de betekenis is vastgelegd in het model en computers begrijpen wat er wordt bedoeld. ProRail zet hiermee de eerste stap op het pad van de Linked Data. Zo neemt ProRail de spoorprofessionals mee in de digitale transitie en werken wij slimmer samen.





# grlc: Eenvoudig en snel API's bouwen voor Linked Data

**Albert Meroño Peñuela**

Vrije Universiteit  
Amsterdam

**Richard Zijdeman**

Internationaal Instituut  
voor Sociale Geschiedenis

**Webapplicaties gebruiken steeds vaker Linked Data. Het verwerken van Linked Data kan moeilijk zijn voor de bouwers van een applicatie. Zelfs een eenvoudige taak als het ophalen van gegevens uit een SPARQL-endpoint om vervolgens te tonen in een applicatie, vereist kennis van Linked Data-standaarden zoals RDF en SPARQL. Web-developers moeten derhalve deze standaarden leren, eer ze Linked Data kunnen integreren in hun applicaties.**

grlc biedt een oplossing voor dit probleem en laat aanbieders van Linked Data hun datasets automatisch, snel en collaboratief aanbieden via een Application Programming Interface (API), die elke webdeveloper begrijpt en makkelijk geïntegreerd kan worden. Om dit mogelijk te maken combineert grrc een sociaal codeplatform (GitHub), de Open API Specification-standaard (OAS) en Linked Data-standaarden (SPARQL, RDF) op een unieke manier.

grlc automatiseert de specificatie van de API en genereert de code die de API koppelt met de Linked Data. grrc maakt gebruik van SPARQL-query's die op het Internet openbaar, of eventueel afgeschermd, zijn geannoteerd en gepubliceerd. Deze SPARQL-query's worden op GitHub opgeslagen, omdat Github gebruiksvriendelijk en voor

iedereen toegankelijk is. grrc maakt op basis van deze SPARQL-query's een beschrijving van de API in de Open API Specification-standaard (OAS). De API-specificatie beschrijft welke functies er kunnen worden aangeroepen, wat de parameters van die functies zijn, en welk type gegevens ze aanleveren. Door de API-specificatie in de OAS-standaard aan te bieden, kunnen applicaties hier automatisch mee werken. Ook kan de automatische API-specificatie gebruikt worden om gebruiksvriendelijke API-documentatie te genereren.

grlc genereert ook de code die de API verbindt met de Linked Data. Als een webapplicatie een functie van de API aanroept, dan zorgt de door grrc gegenereerde code ervoor dat de corresponderende SPARQL-query's 'onder de motorkap' worden uitgevoerd en de resultaten als gestructureerde data worden teruggekoppeld.

Omdat grrc uitgaat van Linked Data-standaarden onderscheidt het zich van veel andere API-generatoren die op het Internet te vinden zijn. Met grrc kun je een goed gestructureerde API bouwen voor elke RDF-gebaseerde gegevensbron, inclusief SPARQL-endpoints, RDF-dumps, html-pagina's met ingebedde RDFa-annotaties en gekoppelde Linked Data Fragments-servers. Deze kunnen allemaal worden gecombineerd in dezelfde API waardoor programmeurs geen maatwerk meer hoeven te maken voor elke dataset.

API's maken het voor developers makkelijker om webapplicaties te bouwen met Linked Data.



Sinds de openbare release in juli 2016 wordt grrc actief ontwikkeld door zes ontwikkelaars, waaronder Albert Meroño-Peñuela (Vrije Universiteit Amsterdam), Rinke Hoekstra (Elsevier) en Carlos Martínez (eScience Center) en een community van 2.400 gebruikers, waaronder DANS, TNO, het eScience Center en de RISIS-, CLARIAH-

en Nanopublications-projecten. Voor deze en toekomstige gebruikers blijft grrc interfaces bieden die het genereren en gebruik van API's bovenop Linked Data mogelijk maken.

De webinterface van grrc wordt ook gebruikt om voorbeelden van SPARQL-queries te documenteren en uit te wisselen. SPARQL-queries moeten nog wel met de hand geschreven worden, en grrc wil hier in de toekomst ook meer ondersteuning bij gaan bieden.



# HortiCube: Linked Data voor de voedsel- en sierteeltindustrie

Jack Verhoosel  
TNO

Johan Bremmer  
Wageningen UR

**Bedrijven in de voedselindustrie willen graag weten waar hun producten terecht komen en welke wensen lokale klanten hebben. Met deze kennis kunnen ze hun productie en omzet verbeteren. Telers en handelaren spelen graag in op trends om hun aanbod af te stemmen op de vraag.**

In de voedselketen en sierteeltketen werken veel verschillende partijen wereldwijd om producten te telen, produceren, distribueren, verkopen en consumeren. Iedere partij houdt zijn eigen informatie bij die maar beperkt zichtbaar is voor andere partijen in de keten.

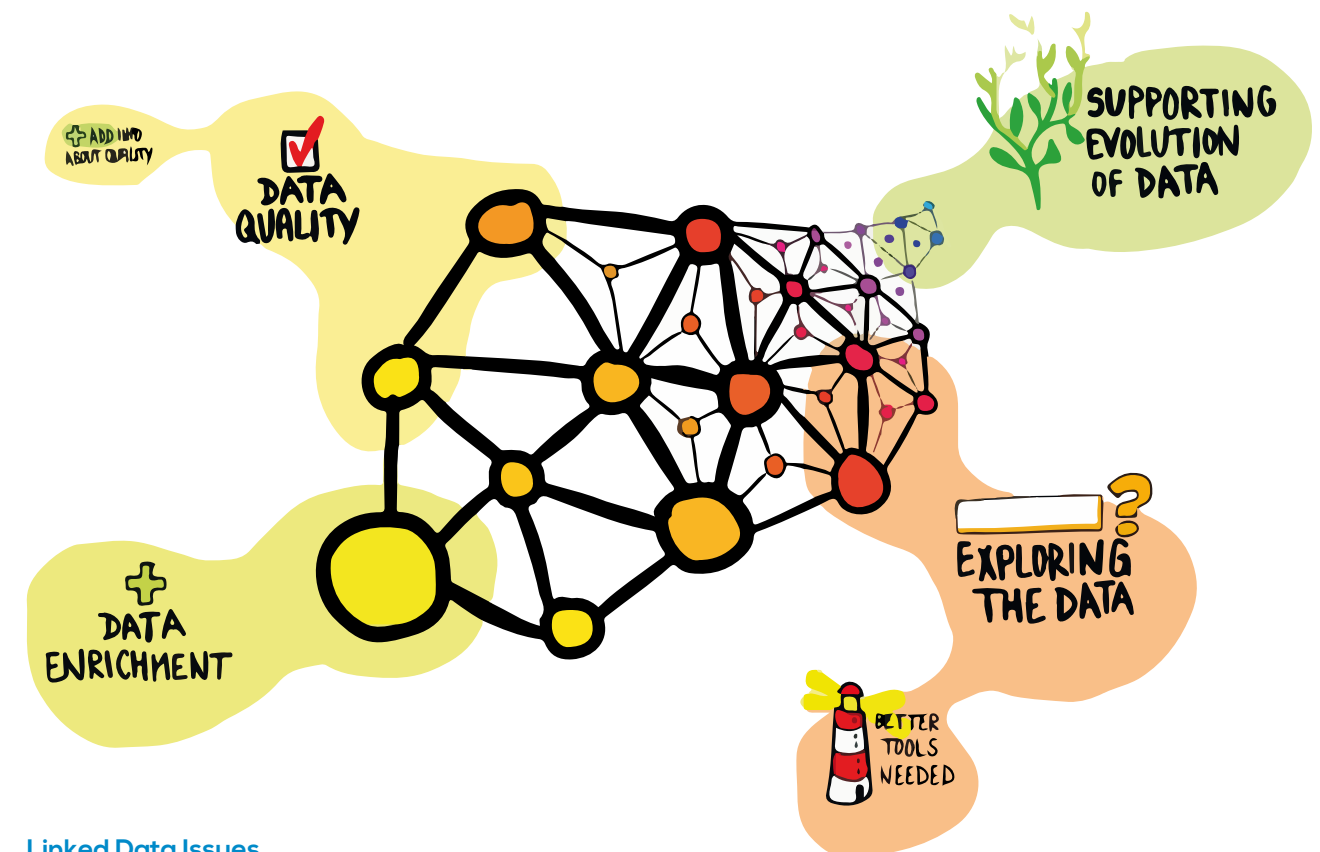
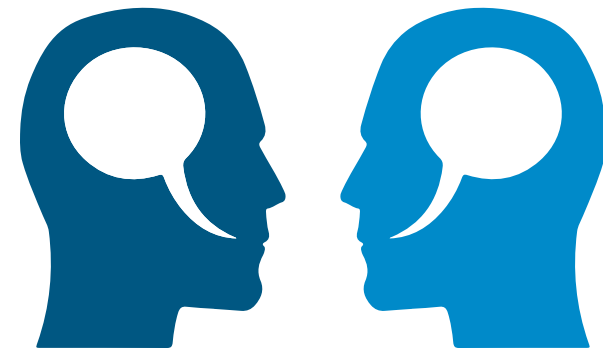
De HortiCube pakt dit probleem aan door toegang te geven tot een combinatie van databronnen in de voedselketen en de sierteeltketen. HortiCube verenigt data van de ketenpartijen zoals productiedata, voorraden, import- en export-transacties, en koppelt deze met openbare data zoals weerinformatie.

Zo verbindt de HortiCube data van de World Apple and Pear Association (WAPA) over productie, productievoorspelling en voorraden van landen in Europa en enkele daarbuiten tussen 2010 en 2018. Ook koppelt de HortiCube de statistieken van de Food and Agriculture Association (FAO) van de Verenigde Naties met productiegegevens van alle voedselproducten van alle landen in de wereld vanaf jaren 70 tot en met 2016.

Linked Data speelt een belangrijke rol in het combineren van deze verspreide gegevenscollecties. De HortiCube gebruikt OWL om een Common Horticultural Model (ontologie) te beschrijven met daarin de termen uit de voedsel- en sierteeltketen zoals 'import', 'export', 'productie', 'voorraad', 'omzet' en 'afzet'. De WAPA- en FAO-databronnen zijn met RDF gepresenteerd als Linked Data-bronnen die ontsloten worden via een SPARQL-endpoint.

De CHM combineert deze databronnen en vertaalt tussen de productnamen en productcodes van de verschillende databronnen. Een rule-reasoner in de

HortiCube helpt de voedselindustrie aan Linked Data om producten te verbeteren en aanbod op vraag af te stemmen.



## Linked Data Issues

HortiCube maakt het mogelijk om een gebruikersvraag op te splitsen in deelvragen die aan de individuele databronnen worden gesteld en die de deelantwoorden weer combineert in een totaalantwoord. De rule-reasoner bevraagt ook gegevensbronnen die niet als Linked Data beschikbaar zijn. Een voorbeeld is de Comtrade-database van de Verenigde Naties die als REST-API bevraagd kan worden. De rule-reasoner kan gebruik maken van 'forward-ruling' en 'backward-ruling' om de vragen te beantwoorden en data te combineren.

De HortiCube is ontwikkeld als prototype binnen het kennisproject BigTU. De HortiCube is als prototype live sinds 2017 en wordt naar verwachting in 2019 in productie genomen. TNO en Wageningen UR werken nu samen met sectorpartijen zoals GroenteFruitHuis, FrugiCom en Q-ray om HortiCube klaar te maken voor productie.

De belangrijkste ervaring met de HortiCube is dat een ontologie helpt om de gebruiker af te schermen van de complexiteit van verspreide gegevensverzamelingen.

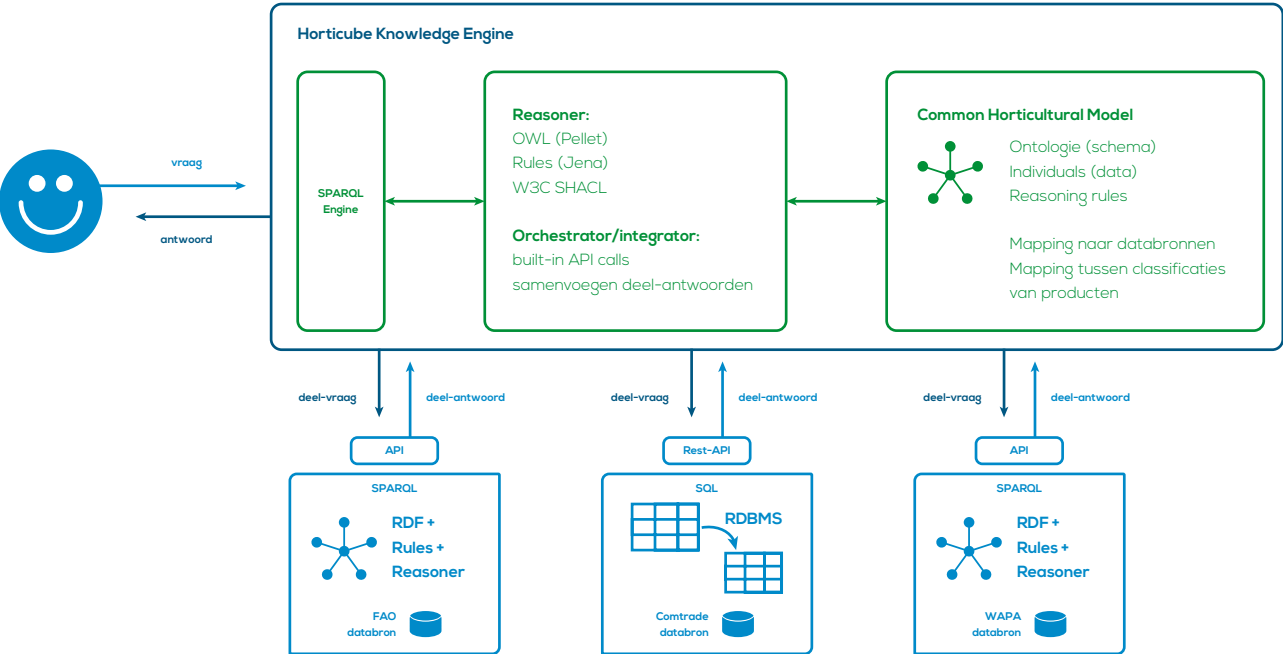


## Lijst met afkortingen

API	Application Programming Interface	NIBG	Nederlands Instituut Beeld & Geluid
BAG	Basisregistratie Adressen en Gebouwen	OAS	Open API Specification
BRT	Basisregistratie Topologie	OTL	Object Type Library
CEO	Culturele Erfgoed Ontologie	OWL	Ontology Web Language
CSV	Comma-Separated Values	OWMS	Overheid.nl Web Metadata Standaard
DTL	Dutch Techcentre for Life sciences	PDOK	Publieke Dienstverlening Op Kaart
ETSI	European Telecommunication Standardisation Institute	PLDN	Platform Linked Data Nederland
FACT	Fair, Accurate, Confidential, Transparent	RCE	Rijksdienst voor het Cultureel Erfgoed
FAIR	Findable, Accessible, Interoperable, Re-usable	RDF	Resource Description Framework
FAO	Food and Agriculture Organisation	RDS	Responsible Data Science
GML	Geography Markup Language	RIO	Registratie Instellingen en Opleidingen
GTAA	Gemeenschappelijke Thesaurus voor Audiovisuele Archieven	SHACL	Shapes Constraint Language
GWSW	Gegevenswoordenboek Stedelijk Water	SKOS	Simple Knowledge Organization System
HTML	HyperText Markup Language	SPARQL	SPARQL Protocol And RDF Query Language
http	HyperText Transfer Protocol	SWRL	Semantic Web Rule Language
IHW	Informatiehuis Water	URL	Universal Resource Locator
IIM	Integraal Informatie Model	VSNU	Vereniging van Samenwerkende Universiteiten
IRI	International Resource Identifier	WAPA	World Apple and Pear Association
JSON	JavaScript Object Notation	WRR	Wetenschappelijke Raad voor het Regeringsbeleid
LOD	Linked Open Data	XML	Extensible Markup Language

Het opstellen van de CHM-ontologie en de vertaling van productclassificaties is intensief werk dat veel tijd kost. Dit moet vanaf het begin met de juiste domeinexperts worden aangepakt.

De HortiCube laat zien dat Linked Data een oplossing biedt waar het samenbrengen van verspreide data-bronnen in één centrale data store geen optie is. Een Linked Data-oplossing vereist wel een goede samenwerking tussen domeinexperts, ontologie-engineers en software-engineers.



# Colofon

## Blijf op de hoogte:

- [platformlinkeddata.nl](http://platformlinkeddata.nl)
- LinkedIN-groep LOD Nederland
- Twitter: @LinkedDataNL
- Nieuwsbrieven: <http://bit.ly/2Pwatkx>

## Redactie

Platform Linked Data Nederland,  
Annemieke Toersen | Forum Standaardisatie  
Han Zuidweg | Forum Standaardisatie

## Illustratie omslag

Lianneh, Zwolle

## Vormgeving

Remwerk, Amersfoort

ISBN: 978-90-365-4675-1



Platform Linked  
Data Nederland

Alle deelnemende partijen die meegeschreven hebben aan deze editie:







Platform Linked  
Data Nederland

[PlatformLinkedData.nl](https://PlatformLinkedData.nl)